Application of a Markov Chain Monte Carlo calibration and uncertainty framework to a process-based integrated nitrogen model (INCA)

B. M. Jackson[1], H. S. Wheater[1], N. McIntyre[1], P. Whitehead[2]

[1] Department of Civil and Environmental Engineering,
Imperial College London,
London SW7 2BU,
United Kingdom.
Tel: +44 20 7594 6122
Fax: +44 20 7594 6124
Email: b.m.jackson@imperial.ac.uk

[2] Department of Geography,
The University of Reading,
Reading RG6 6AB, UK
Telephone: +44 (0)118 378 8733
Fax: +44 (0)118 975 5865

ABSTRACT

An integrated nitrogen model has been developed to investigate the fate and distribution of nitrogen in the aquatic and terrestrial environment. This Integrated Nitrogen in CAtchments model (INCA) simulates flow, nitrate and ammonium and tracks the temporal variations in hydrological flow paths and nitrogen mass operating in both land and river phases. Dilution, natural decay and biochemical transformation processes are included in the model as well as interactions with plant biomass. It is semi-distributed to account for spatial variations in land use, human impacts, effluent discharges and varying deposition levels, and produces daily estimates of the stream water flow and nitrate and ammonium concentrations, in addition to estimates of annual, land-use specific, N fluxes.

While the model has been successfully applied to a range of catchments in the U.K. and Europe, little work on parameter sensitivity and identifiability has been carried out on it to date. A detailed study of these issues should aid the successful calibration of further catchment applications by highlighting the most significant parameters and allowing informed decisions as to the areas in which experimental resources and measurements should be allocated.

It is also necessary to provide measures of the uncertainty present due to measurement errors in the inputs, parametric uncertainty, and issues related to model conceptualisation. INCA, describing a set of complicated environmental processes with dependencies on both space and time, is necessarily a simplified representation of the phenomena being studied. This imposes a limit upon one's confidence in its responses or outputs, regardless of the accuracy of any input information. The input itself is subject to many sources of uncertainty, including measurement errors, absence of information and incomplete understanding of underlying driving forces and mechanisms. Adequate spatial representation is particularly difficult, due to the intrinsic variability present within the environment, such as the continuous variation in soil properties and nitrogen inputs over space, and the difficulty of characterising

properties in the subsurface. Therefore, spatial information is likely to be severely limited in any application of the model.

To address these considerations, a collection of Monte Carlo routines within a subjective probability framework has been developed for use with the INCA model. Markov chain Monte Carlo methods (using Metropolis Hastings formulae) are used to sample parametric and uncertain quantities. The framework permits both parametric and model structural uncertainty to be interrogated, and allows effective calibration and confidence predictions through optimisation of model inputs to fit observations or other criteria, with explicit consideration of effects of data uncertainty.

MCMC methods possess the general virtue of simulation methods, with information regarding parametric probability distributions easily collected along with optimal parameter sets. However, other sampling methods generally fail when the posterior involves many variables or is otherwise intractable. Markov chain methods are capable of sampling from posterior distributions of arbitrary complexity, through the Metropolis Hastings algorithm, which provides simple conditions under which the chain will equilibrate to the required distribution. Since such methods sit naturally within a subjective probability framework, they are also capable of quantifying distortions produced on the outputs by noise. Such a capability is indispensable for rigorous analysis of an environmental model such as INCA, as the input is subject to extreme uncertainty.

Markov chains are constructed such that their equilibrium distribution is equal to the posterior distribution of interest, and each state is visited the required number of times to satisfy the conditional distribution of the parameters given the data. This is achieved through satisfying appropriate conditions of reversibility (detailed balance) and ergodicity. By giving the microscopic dynamics of the Markov chain the (unnormalised) input distribution is implicitly fixed. This allows the treatment of problems that are too complex for an explicitly specified input distribution, such as identifying the posterior uncertainty of the INCA parameters given optimality constraints, or efficiently reconstructing the inputs from output distributions.

The equilibrium, or *posterior* distribution, is obtained via Bayes' rule. This equation describes the current knowledge regarding parameter distributions, given initial knowledge and information from prior runs, and allows potential for converging upon the "true" input distributions through incorporation of learnt information. The draws from the Markov chain are accomplished through variants of the pleasingly simple Metropolis Hastings formulae, involving proposals of candidate values through a proposal function and rejection/acceptance steps. This proposal function is constructed such that it implicitly defines the required conditional distributions, along with satisfying the necessary Markov chain conditions. For the INCA applications described here, distributions have been defined as conditional upon aspects of the model response and optimality criteria imposed upon the analysis. Proposal functions have also been constructed to allow sampling to be efficiently weighted towards subsets of the distribution where this is desirable.

The performance of the modelling framework is illustrated with data from the Kennet catchment in southern England. This is a groundwater-dominated catchment draining an area of 1164 km^2, with a chalk aquifer supplying approximately 95% of its water.

As it has been a focus of a variety of water quality and ecological concerns, there is a relatively large amount of data available to compare model response against. To understand the characteristics of both overall uncertainty and particular parametric sensitivities in INCA, the effect of changes in the parameters and inputs are examined using the Markov chain sampling described above. *Response surfaces*, in this case distributions of input parameters against single-valued measures of performance (derived from the output parameters and optimality criteria), are examined, and the biases caused by differing optima considered. The influence of such biases on subsequent decisions regarding parameter sensitivities and “optimal” parameter sets is examined. The efficiencies of differing Metropolis proposal functions applied to sample both the “minima” and entirety of a given response surface are also investigated.

To address questions regarding the appropriateness of the model structure and propose efficient calibration strategies, results are applied to identify components of the model structure that appear most significant in the simulation of nitrogen dynamics in river systems, aspects that appear redundant, and the most pertinent data for model/process identification. The relative importance of differing measurands over space is also examined, and suggestions for subsequent allocation of experimental resources made. Residual model uncertainty is translated into prediction confidence limits for management purposes.

The results demonstrate the power of Markov chain Monte Carlo methods to quantitatively examine the inter-relationship between model structure, parameter identifiability and data support, and also provide an efficient means of addressing the problem of calibration given large parameter sets and the presence of measurement error and other uncertainties.